PB86-210465

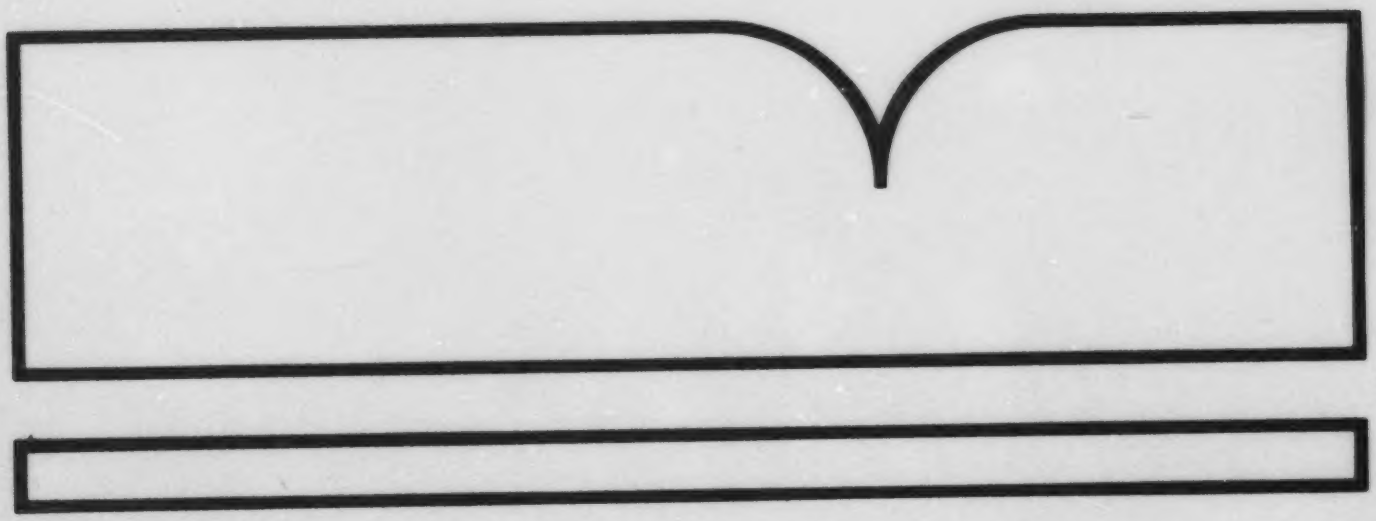
Controlling Outliers in Contingent Valuation Experiments

Yale Univ., New Haven, CT

Prepared for

Environmental Protection Agency, Washington, DC

Apr 86



U.S. Department of Commerce
National Technical Information Service
NTIS

United States
Environmental Protection
Agency

Office of Policy
Planning and Evaluation
Washington DC 20460

EPA 230-06-86-015
January 1986

PB86-210465

EPA

Controlling Outliers in Contingent Valuation Experiments

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

30272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO. EPA 230-06-86-015	2.	3. Recipient's Accession No. PB86 210465/AS
4. Title and Subtitle: Controlling Outliers in Contingent Valuation Experiments			5. Report Date: April 1986 6.
7. Author(s): Robert Mendelsohn			8. Performing Organization Rept. No.
9. Performing Organization Name and Address: School of Forestry and Environmental Studies Yale University New Haven, CT 06511			10. Project/Task/Work Unit No. 11. Contract(C) or Grant(G) No. (C) 68-01-6596 (G)
12. Sponsoring Organization Name and Address: Office of Policy Analysis, PM-221 U.S. Environmental Protection Agency 401 M Street, S.W. Washington, D.C. 20460			13. Type of Report & Period Covered: Final 14.

15. Supplementary Notes

16. Abstract (Limit: 200 words)

There is no established protocol for the analysis of data collected in contingent valuation surveys. This has led to seemingly arbitrary removal of some outlier observations, which leads to questions about the representativeness of the results reported in the literature. This report critiques the methods researchers have used to analyze the data from contingent valuation surveys and suggests procedures for future analyses. Three procedures regress the responses on a set of economic and demographic variables (e.g., income, age, sex, prices). Then if the unweighted residuals appear to be normally distributed, analysis should proceed. The first process would drop outliers (iteratively) that are more than (say) six standard deviations from the mean response. The second process would weight observations close to the regression line more heavily than those further away, so that those with true extreme preferences are not eliminated. The third procedure would incorporate known independent sources of bias in the regression. A forth approach would examine the median response rather the mean; this would be more like a voting outcome but ignores extreme preferences. Testing of competing approaches is recommended.

17. Document Analysis a. Descriptors
contingent valuation analysis
protest bids
outlier bids

b. Identifiers/Open-Ended Terms
contingent valuation surveys, analysis of willingness-to-pay survey responses,
data analysis of CV surveys,
accounting for skewed distributions

c. COSATI Field/Group

18. Availability Statement: No restriction on distribution. Available from National Technical Information Service, Springfield, VA 22161	19. Security Class (This Report): Unclassified	21. No. of Pages: 26
	20. Security Class (This Page): Unclassified	22. Price

(See ANSI-Z39.18) See Instructions on Reverse

OPTIONAL FORM 272 (4-77)
(Formerly NTIS-35)
Department of Commerce

CONTROLLING OUTLIERS IN CONTINGENT VALUATION EXPERIMENTS[a]

ROBERT MENDELSOHN

YALE SCHOOL OF FORESTRY AND ENVIRONMENTAL STUDIES

205 PROSPECT STREET

NEW HAVEN,CONNECTICUT 06511

APRIL,1986

[a] This research is supported by the Office of Policy Analysis of the U.S. Environmental Protection Agency. Many thanks go to Ann Fisher for her thoughtful and careful guidance throughout the project. Comments on an earlier draft from Jon Harford, Robin Gregory, Kerry Smith, and Reed Johnson have also been very helpful.

I. INTRODUCTION

The use of contingent valuation as a tool to measure the value of environmental amenities has been of growing interest to both economists and policy makers. From the initial work of Davis [1963] and Hammack and Brown [1974] to the more recent work by Brookshire et al. [1982], Desvousges et al. [1983], and Tolley and Randall [1984], there have been multiple applications of contingent valuation to many valuation questions of great policy importance.

Although contingent valuation approaches have gained a certain degree of acceptance from resource economists (they are recommended by the Water Resources Council, for example), there remain methodological questions about possible biases and inaccuracies. For example, researchers have noted the possibility of interviewer bias, hypothetical bias, payment vehicle effects, starting point bias, information bias, and strategic bias (see Schulze et al.[1981], Mitchell and Carson [1982], or Cummings et al.[1984] for reviews of these effects). Several experiments have examined the extent to which these effects exist and whether any existing problems can be mitigated. Most efforts to mitigate these potential problems have focused upon survey design and sampling. This emphasis on survey design is probably appropriate since a badly designed survey instrument can completely obscure useful information.

However, little attention has been given to the data analysis of existing surveys (see Brookshire et al. [1982] and Desvousges

et al. [1983] for two notable exceptions). This has led to three potential deficiencies in the literature. First, the literature has failed to develop a protocol for proper analysis and reporting of results. Second, the seemingly arbitrary removal of observations in some CV studies raises questions about the representativeness of reported results. Third, there remains the potential that more sophisticated analyses of obtained data may be able to mitigate some of the alleged problems with the entire CV approach.

The purpose of this study is to review methods researchers have used to control outliers in contingent valuation (CV) surveys and to recommend new approaches to this problem. In the process, we will identify the pitfalls of certain approaches and hopefully standardize some methods of data analysis for the CV literature.

Contingent valuation researchers tend to apply only simple statistical techniques to their survey data. For example, many researchers merely compute and report the means of their responses. This simplicity is, in fact, one of the attractions of the approach. The results of a CV experiment can be calculated by individuals with relatively little technical training and the entire approach is easy to understand and thus easy to communicate.

The central problem we wish to examine is whether or not to weight all responses equally. There are three issues relevant to this discussion. First, the distribution of CV responses tends

to be highly skewed leading researchers to question whether or not they should truncate the distribution above a certain value. Second, there tend to be many zero responses to CV questionnaires which raises the possibility that responses are truncated at zero or that there are many protest or nonparticipating responses. Third, there is the possibility that some participants are not as involved in the hypothetical decision as desired and are giving erratic or biased responses to questions. To the extent that some responses can be identified as inaccurate or biased, it would be desirable to remove or at least deemphasize them in the analysis.

The next section of this paper reviews existing approaches of data analysis that control for less desired observations. The strengths and shortcomings of each approach are discussed. Some alternative methods to control for outliers and biased responses are suggested in Section III. Finally, Section IV summarizes the results and suggests a series of tests to measure how well each control methodology performs.

II. HISTORIC CV DATA ANALYSIS

Because contingent valuation questions typically are designed to ask precisely what is desired to be known, the mean of the responses is often the sole reported statistic. Of course, the standard deviation or variance of responses should also be reported so that readers can verify if the responses are statistically different from zero or some other nonzero response of interest.

As is inevitable with any large data collection effort, there are bound to be measurement errors associated with the data. "Cleaning" the data, trying to reduce measurement error, is a normal part of any data collection process. Obviously, straightforward corrections such as removing coding errors and misprints can only lead to more accurate responses. It is less clear, however, how one ought to identify and then deal with problem interviews. The contingent valuation literature regularly eliminates problem responses from collected samples. The critical issue is how are the removed observations identified. Depending upon the editing procedure, removal of problem observations will either enhance the accuracy of the results or cast a shadow upon the entire study.

One problem with CV responses is that they are concentrated at zero. Early practitioners of the CV approach were concerned that the zero responses reflected a lack of cooperation from the respondent and labelled these votes as "protest bids". Protest bids at zero tend to bias down survey means. Some researchers have responded by removing zero bids.

A second problem with CV responses is that they are skewed with many responses in the upper tail. These responses can be so far above the mean that they have enormous impacts upon the reported means. The outliers also clearly contribute to large standard deviations which tend to reduce the reliability of the reported results. Thus, several researchers have eliminated the upper tails of the responses from their surveys. Unfortunately,

researchers have not always been forthcoming about either their methodology for editing the data nor precisely what responses were removed. The removal of zero responses and the upper tails can give a misleading impression of results. The absence of a reliable standardized analytical methodology in CV experiments raises serious questions.

Removing data improperly reduces the value of a study. In particular, if researchers tend to remove observations which differ from their priors, they will bias the experiment. In the extreme, the likelihood that any prior is accepted will depend solely on the extent of editing. Stated strongly, a paper which uses an arbitrary approach to remove data (or equivalently from the reader`s perspective does not report how it edited the data) and then reports only the edited data should always be viewed suspiciously. Thus, as a methodological point, it is essential that all data removal be done objectively and systematically.

Not all studies have employed subjective criteria for data removal; several have adopted objective criteria. One approach is simply to establish an absolute maximum acceptable response. The maximum could be established before the sample is collected or during the analysis. If the acceptable limits are established during the analysis, the analyst once again has the power to produce whatever result he desires. If the actual mean is judged "too high", the analyst merely can reduce the maximum allowed response. Clearly such an endogenous process has all the pitfalls of subjective editing mentioned above. Even if the

maximum is established in advance of data collection, there remains the possibility that the maximum can bias the results. For example, suppose one believed that the value of hunting was $50 a year per hunter and not the $100 value asserted by another study. Eliminating all responses above $500 may increase the chance that the mean is closer to $50 rather than $100. Lowering the maximum acceptable limit to $200 may further increase this possibility.

Having both a maximum and a minimum permissable response can control the outcome even more. Although the application of the editing rule with exogenous boundaries is systematic, the construction of maximums and minimums tends to be arbitrary. Thus a researcher who is concerned with measuring the high end of true tastes may want to establish high maximums, whereas another researcher more concerned with upward bias from aberrant high responses may establish a low maximum. Even with the same data, these two researchers could come up with significantly different results. The establishment of maximums can truncate the true distribution and thus both bias the reported means and underestimate the true variances. As long as such limits cannot be supported with corroborating evidence, the use of absolute maximum and minimum cutoffs cannot be recommended for CV experiments.

A second objective approach (recommended in Brookshire et al. [1982]), is to remove all data which are eight or more standard deviations from the sample mean. In some circumstances, this proposal is quite reasonable. For example, suppose we

believe there is a true value with a small amount of noise around it. If this noise is distributed normally, then the probability that an observation is 8 standard deviations from the mean response is very small. The outlier can be discarded because it is not likely to have come from the true distribution.

Although the standard deviation approach avoids the potential bias problem of absolute limits, it is not without problems. First, the actual distribution of responses to CV questionnaires tends not to be normal, but truncated at zero and highly skewed. This makes it unclear whether observations many standard deviations from the mean are from the true distribution. There is a great deal of evidence in demand theory that tastes will be systematically different across people with different incomes, ages, sex, etc. The standard deviation approach ignores these sytematic components by assuming there is one true response. One such systematic variable could be membership in a mountain climbing club. Within the club the value given to a specific climbing site may be fairly homogeneous. However, if the club members formed a small component of a community, their values may be very far from the community average. For example, suppose a general population survey was conducted with a question asking each respondent the value they would place upon access to a sheer mountain cliff. It is not hard to imagine a near zero response from the 99.99% of the population not in the mountaineering club. However, the one in ten thousand people who are club members may be willing to spend considerable sums to climb that cliff. An eight standard deviation approach may well eliminate all the

mountain climbers from the included sample.

Second, there are technical issues concerning which standard deviation to apply. As more and more outliers are removed, the standard deviation of the remaining observations tightens thus making yet additional observations appear to be outliers. Presumably, this tighter standard deviation includes only "true observations" that should be used. Thus, the method must be done iteratively until no more observations drop out. Also, the choice of 8 standard deviations or 10 standard deviations is largely arbitrary. If samples were symmetric, this arbitrariness wouldn't affect reported means although it obviously would affect the variance in responses. However, since samples tend not to be symmetric but rather are highly skewed, the choice of range for the sample will affect the reported mean. Thus, the choice of outer boundary through standard deviations potentially has the same effect as the choice of absolute maximums. To achieve a lower reported result, one could merely eliminate all observations greater than six rather than eight standard deviations away.

Mendelsohn, Moore, and Whitelaw [1984] propose a third mechanism to deal with outliers. Instead of relying upon means as a summary statistic, report median responses. One justification of the median approach suggested by Phil Meyer is that the CV approach mirrors the attitudes of voters. In a political system with majority rule, it is the value of the median voter which matters. Thus, policy makers ought to be interested

in the median response to CV questionnaires.

A second attractive quality of the median is that it is robust against individual bias. That is, a single individual can have only a minor impact upon the median response to a survey questionnaire. In contrast, any individual could have a dramatic impact upon the mean by vastly overstating true values. For example, suppose a population tended to value a particular wilderness area as worth $50 per year. An individual who arbitrarily or strategically responded with $5000 per year could seriously alter the mean. For example, the above could occur through a misunderstanding of the question. With willingness to accept, some people may actually be positioning themselves for a handsome payment. Finally, a respondent may be trying to raise the mean response on purpose. Several studies of CV questionnaires have explicitly looked for this last effect,strategic bias, and report that strategic responses are not widespread. However, even if only five percent of a sample respond strategically, they can still significantly alter the mean response. The infrequency of strategic responses, however, suggests that median responses would not be affected by this behavior. Thus, one potential advantage of median responses is that they eliminate the influence of rare but extreme biased responses.

There are serious problems with median analysis, however. First, the median voter hypothesis is a rather simplistic version of politics which totally ignores strength of preference and interest groups. Highly motivated minority groups who may in

fact have disproportionate influence on the relevant issues may be totally ignored by the median statistic.

Using the median rather than the mean suffers from the same problem as removing distant outliers. The median, by ignoring the responses in the tail of the distribution, may be eliminating key minorities and thus be further from the truth rather than the mean.

Second, although it is straightforward to calculate medians, it is not obvious how to use the resulting estimate for hypothesis testing. Unlike the extensive literature on means, there are far fewer published statistics and distributions which help sort out the differences in medians between any two populations. Although there are statistics which could be used as indications of whether any two medians are alike or not, these statistics are not as formal as hypothesis tests using means.

One approximation of whether any two medians are alike is to treat the median as though it were the mean and use traditional hypothesis tests. Using the standard deviation, S_i, computed on the mean, the t statistic between two medians would be:

$$t_r = (X - Y) \left[n * m (n-1) (m-1) / (n+m) \right]^{.5} / (n * S_x + m * S_y)^{.5}$$

where n and m are the number of observations of each sample and X and Y are the medians. In one sense, this adjusted t statistic overstates the true reliability of the median because the true standard deviation using the actual median, not the mean, is

higher. In another sense, this adjusted t statistic understates the true reliability of the median because the median is not affected by extreme responses in the tail of the distribution and yet the standard deviation is quite sensitive to these values. Although this overall test statistic is bound to be too high or too low in certain cases, these two effects tend to offset each other so that the adjusted t statistic is probably a reasonable approximation.

Another suggestion by Tukey [1970] is to divide the data into eighths. Calculate the cutoff value at the end of each section. The median value is now surrounded by four higher and lower cutoff values. Any two medians which fall within the first set of cutoff values are "close". As the medians fall into more and more distant cutoff values, the samples are more distinct.

For example, suppose the 0%, 12%, 25%, 37%, 50%, 62%, 75%, 87%, and 100% responses to a survey were respectively: 0, 10, 15, 25, 40, 75, 125, 370, and 4230. In this case, 40 is the median response. If the median of another sample fell between 25 and 75, the two samples would be considered close. On the otherhand, if the other sample fell outside of 10 and 370, the samples would be considered different.

In summary, the advantage of using medians is that they are insensitive to the extreme tails of a distribution. Medians consequently may be more robust than means. One disadvantage is that the extreme tails of distributions may sometime contain important values. A second important disadvantage is that one is

giving up a battery of well defined statistical tests when moving to medians.

Randall, Ives, and Eastman [1974] propose a fourth objective approach for removing problem observations. For all zero responses to contingent valuation, they recommend a follow-up set of questions to determine whether the zero valuation was really a protest vote (downwardly biased) or a true response. Only respondents who subsequently state that the zero values are protest votes are removed. This approach is attractive in that it uses an exogenous factor (in this case an independent question) as a basis to remove the observation. The removal of zero protest votes in this fashion can only improve the accuracy and consistency of the data. Of course, the frequency of removed protest votes is a useful measure in determining the reliability of a question. Even if protest votes should be removed, their frequency is still a useful statistic.

As Desvousges et al. [1983] note, however, the Randall et al. [1974] approach does not go far enough. Only zero value responses are removed because of a tendency for protest. However, individuals objecting to a specific question need not necessarily answer zero as their only response. They might respond with the starting bid and thus quickly end further discussion of this issue, or they could bid a preposterous amount as a form of protest. Thus, given that an independent question is going to be asked to measure whether or not a previous response was actually a protest, it should probably be applied to

the entire sample. Applying the protest bid test and removal scheme to just zero bids will only detect negative protest votes. If there are just as many bids biased upwards as downwards, correcting just the zero protest bids will tend to bias the corrected mean responses upwards.

Desvousges, Smith and McGivney [1983] (DSM) suggest an alternative method to remove outliers. Recognizing that the true underlying bid is determined by an underlying demand model, they suggest estimating a reduced form demand model with the entire data set. For example, one could regress the contingent valuation responses upon a set of demographic variables such as age, sex, education, or income and a set of price variables such as distance from the site or entrance fees. Demand theory suggests that some of these variables ought to affect the true contingent valuation response. Thus, these variables should explain some of the observed variation in responses across individuals.

DSM then suggest applying a Belsley, Kuh, and Welsch [1980] regression diagnostic to the estimated equation. More specifically, they recommend applying a sensitivity statistic called DFBETA. DFBETA is defined:

$$\text{DFBETA} = \frac{(X^T X)^{-1} x_i e_i}{1 - x_i (X^T X)^{-1} x_i^T} = b - b(i).$$

DFBETA calculates the difference between an estimated coefficient with the whole sample (b) and the same coefficient estimated upon the sample minus the ith observation (b(i)). This statistic is intended to measure the importance of a specific observation upon an estimated coefficient. Thus, if one were doing a study looking for the value of a specific coefficient, the DFBETA coefficient would provide information about the robustness of the estimate and also which observations were most important in determining the estimated value of the coefficient. For example, if one were concerned that a specific coefficient was being affected by variables not included in the regression, one could examine the observations most critical to the coefficient and look for missing variables which are consistent qualities of these observations.

Because income is the only variable which apriori was expected to affect the CV response, DSM recommend focusing upon this coefficient in the regression analysis. They suggest estimating the percentage change in DFBETA for the income coefficient. This percentage change variable is labelled the BKW statistic and it is defined as:

$$BKW = DFBETA/b = [b-b(i)]/b$$

where in this case, the coefficient b is specifically the coefficient on income in the CV regression. The authors then recommend removing all observations with a BKW above 30%.

In their Monongahela study, Desvousges et. al. [1983] apply

this methodology to their CV experiment. Using the 30% criteria on the income coefficient, the authors remove 13 percent of their sample (32/243). The mean response of payments to avoid losing further water quality on the Monongahela River both with and without outliers is:

	With Outliers	Without Outliers
$125 Start	68	57
$25 Start	28	29
Payment card	55	51
Direct	30	25

With the outliers removed, the authors compare the t statistic between all 6 possible combinations. DSM find the two way comparisons all significantly different except for the comparison between the $125 and $25 starting points and the payment card approach. However, including the outliers and redoing this experiment, we find the same set of results.

It may seem somewhat surprising that one could remove 13% of the observations, all outliers, and yet still have no perceptible effect on the results. However, closer inspection of the BKW statistic reveals why it may not affect the CV average. The BKW statistic was invented to examine the impact of specific observations on the computed regression coefficients. The BKW statistic takes the error in the regression equation and weights it by the deviation of the independent variable from the mean. Thus, if a respondent with the mean level of income gives a large error, the BKW statistic gives it little weight. On the

otherhand, if a person with above or below average income has a large error, the BKW statistic gives it great weight. Looking at Figure I, three observations are illustrated. Each illustrates a constant positive error above the regression line. The BKW statistic will identify x_h and x_l as important because these two error terms have the greatest impact on the regression line. Although x_m distorts the mean CV response just as much as x_h and x_l, the BKW statistic ignores this error because it has no impact on the coefficient for income. Thus, if we were specifically interested in how the CV responses change with income, the removal of the 13% of the sample which most affects this coefficient could have a large impact.

However, the purpose of most CV experiments is to estimate the mean response to a question not the correlation between responses and income. The DFBETA coefficient is appropriate for this latter question but not the former. It is not clear that the DFBETA coefficient should be recommended for most CV analyses.

III. ALTERNATIVE METHODS

In this section, some alternative approaches are recommended for analyzing the results of CV experiments. In each of these approaches, CV responses are regressed upon a set of independent variables. The first technique uses the regression equation to eliminate outliers. The second approach uses the residuals of the regression equation to weight observations. The third approach uses the regression equation to search for measured

biases.

The attractiveness of the Desvousges, Smith, and McGivney [1983] approach is that it takes account of an underlying model which explains true variations in CV responses. The following three alternatives builds upon the DSM approach. In each, the CV responses are regressed upon a set of economic and demographic explanatory variables such as income, age, sex, and prices.

In the first approach, we intend to drop outliers. Thus, all observations greater than, for example, six standard deviations from the expected response would be dropped. The regression could then be repeated, the standard error of the residuals could be reestimated, and new outliers dropped. The process ends when there are no more outliers to eliminate.

The advantage of using the regression model in the elimination of outliers is that true variations in response can be accounted for. In the mountain climbing example, including mountain climbing experience as an independent variable, would permit the regression equation to predict the minority with strong views on this sport. The true variation in opinions on mountain climbing would be part of the regression model not part of the error term. The regression model thus potentially overcomes some of the most serious problems with taking just the deviations from the mean response as the criterion. Similarly, the focus upon the residual rather than the DFBETA statistic keeps our attention upon how an error term affects the mean CV response rather than upon how it affects the relationship between

CV and income. Thus, it is the elimination of unexplainable outliers that is the focus of the analysis.

In the second approach to handle outliers, all observations are weighted by how close the observation lies to the estimated regression equation. Observations close to the regression equation would be given more weight than observations far from predicted values. The mean of these weighted CV responses and the standard deviation of these weighted values would then be the reported values. For example, one could develop a standardized residual where the residual is expressed in terms of standard deviations from the predicted value. The observations could then be weighted by the inverse square of the standardized residual. Thus, observations near the regression line (for example, .01 standard deviations from the line) would be given a weight of 100 whereas observations two standard deviations from the regression line would be given a weight of .25. The advantage of this weighting scheme is that it recognizes that deviations from the line are a continuous variable rather than an arbitrary discrete cutoff. All observations remain in the sample, but well behaved observations are given relatively more importance than less well behaved points.

A third potential regression approach is to incorporate known independent sources of bias into the analysis. For example, the same CV question could be asked in more than one way. The consistency of response could then be used as an independent variable. Similarly, with iterative bidding , one could include

the starting payment level as a possible source of bias. Similarly, with payment cards, the range and arrangement of payments suggested on the card could be the source of some bias. The specific interviewer could be a potential source of bias. Intentional variations in information given prior to the question could serve as a bias marker variable. The payment vehicle itself could be a bias marker. In all these cases, if the independent variable is measured, it may be possible to adjust for the existing bias through analysis of the resulting regression coefficient.

For example, Jon Harford recommends using the following model to adjust for starting point bias. Defining S to be the starting point, suppose the individual's observed response is V and it relates to the true response X in the following way:

$$V = X + b(S-X) + e \quad ,$$

where e is a random error. The model posits that the starting point leads people away from their true values through a linear distortion which depends upon the size of the coefficient b. To obtain the true expected value of X, one must take the expected value of the above expression and solve for X:

$$X = (V-bS)/ (1-b).$$

Including the independent measure of bias into the regression equation, estimating the equation, and then recomputing the results can lead to a bias-free measure of the CV response.

Similar results could be obtained for the other sources of

bias provided it is understood what is the truth and what is a source of bias. That is, the regression equation can detect the effect of having more or less information, different payment vehicles, or more consistency between responses. Sometimes, these sources of bias are easier to correct than others. For example, consistent responses across questions is generally desirable. Observations which are shown to be consistent could be given greater weight. More difficult are the biases where the true response is not known. For example, which payment vehicle elicits the true response versus the distorted response? If it is not clear, one could use the regression coefficients to estimate different answers depending upon the vehicle. Thus, the coefficients could be used to provide a range of plausible responses or a sensitivity test. The inclusion of independent variables which might measure bias in the overall regression explaining CV responses is yet another way that analysis might help improve the overall quality of the CV approach.

IV. CONCLUSION

In section II of this report, several methodologies to "clean" contingent valuation responses are reviewed. The subjective removal of data which conflicts with analyst`s priors is bad science. Even objective methods to control outliers must be adopted cautiously. Several methods suggested in the contingent valuation literature such as absolute minimums and maximums, medians, and standard deviation cutoffs have serious deficiencies. This paper builds upon a suggestion by Desvousges,

Smith, and McGivney [1983] to use regression analysis to control outliers. Using regression equations to mitigate bias, remove outliers, or weight observations are all suggested in this paper.

In order to develop an analytical methodology for CV experiments, it is important to test these competing approaches. Each of these methods need to be explored using samples with potential verification abilities. For example, one could use a survey which asks both willingness to pay and willingness to sell questions for the same good. Comparisons betweeen responses to the two questions provide a measure of how well each methodology removes problem responses. The mean response of the original untransformed sample could then be compared with the response after each of the control methodologies. One could test whether the data transformations can help remove the unwanted difference between WTP and WTS.

Similarly, a comparison between contingent valuation and a revealed preference model such as a travel cost or hedonic model, would be attractive. The adjusted and unadjusted CV response could be compared to the travel cost response. One could potentially test starting point, interviewer, and payment vehicle bias with such a data set. Through empirical tests such as these, the value of using appropriate objective methods to analyze CV responses could be explored. Not only may these techniques improve the accuracy of CV responses but they may also help solve some of the bias problems often cited in the literature.

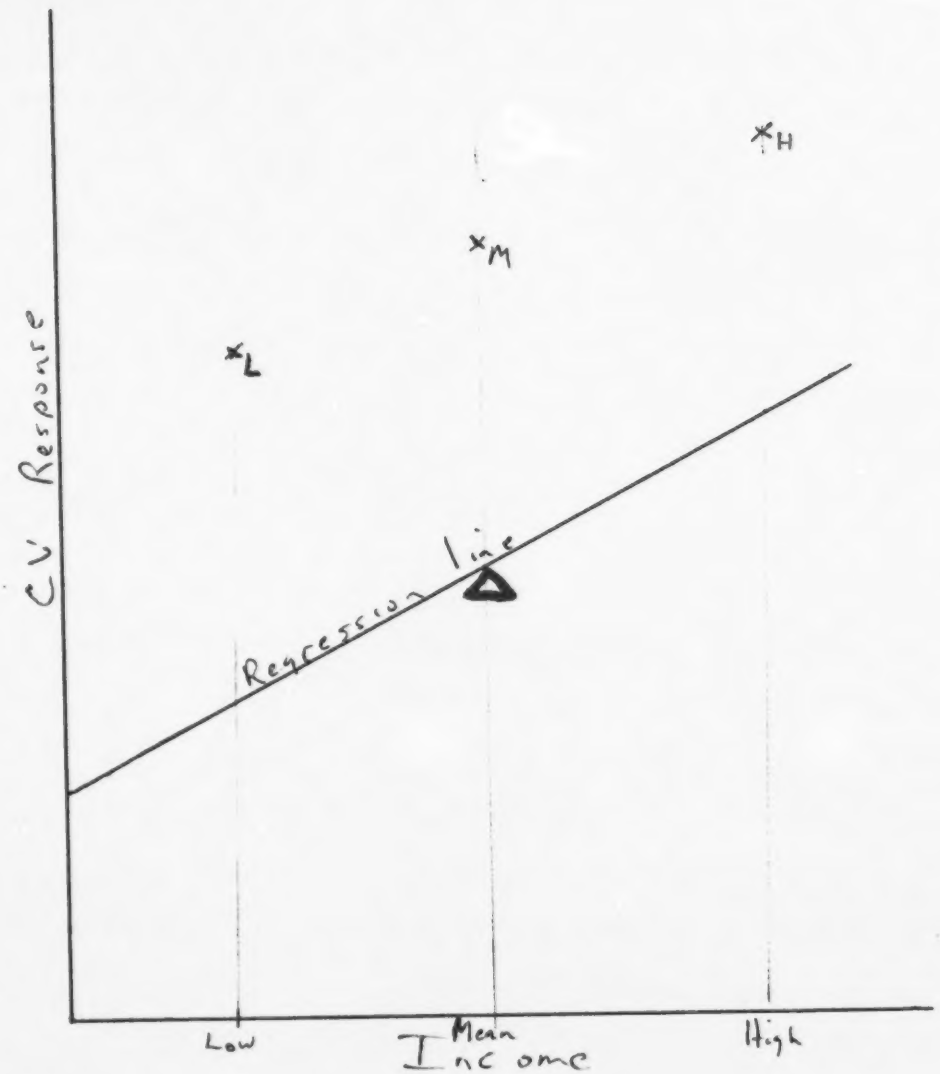


Figure I

BIBLIOGRAPHY

Belsley,D.,E. Kuh, and R. Welsh,Regression Diagnostics New York: John Wiley and Sons, 1980.

Brookshire,D. M. Thayer, W. Schulze, and R. D'Arge, "Valuing Public Goods: A Comparison of Survey and Hedonic Approaches". American Economic Review 72 (March,1982) 165-77.

Cummings,R.,D. Brookshire, W. Schulze, and D. Coursey, " Valuing Environmental Goods: A State of the Arts Assessment of the Contingent Valuation Method" US EPA, Washington, D.C. 1984.

Davis,Robert, "Recreation Planning as an Economic Problem" Natural Resources Journal 3 (Oct.1963) 239-49.

Desvousges,W., V.K. Smith, and M. McGivney, "A Comparison of Alternative Approaches for Estimating Recreation and Related Benefits of Water Quality Improvements" US EPA, WAshington D.C. EPA-230-05-83-001, 1983.

Hammack,J. and Gardner Brown, Waterfowl and Wetlands: Toward Bioeconomic Analysis. Baltimore: Resources for the Future,1974.

Mendelsohn,R. Terry Moore, and Ed Whitelaw, "An Economic Evaluation of Fish Loss in the Swan River Drainage" Montana Fish Game and Parks, 1984.

Mitchell,Robert and R. Carson," Commetn on Option Value: Empirical Evidence From A Case Study of Recreation and Water Quality" Resources For the Future, WAshington D.C. 1982.

Randall,Alan, B. Ives, and C. Eastman, "Bidding Games for Valuation of Aesthetic Environmental Improvements" JOurnal of Environmental Economics and Management 1 (1974) 132-49.

Schulze,William, Ralph D'Arge, and David Brookshire, "Valuing Environmental Commodities, Some Recent Experiments",Land Economics 57 (May,1981) 151-73.

Tolley,George and Alan Randall "Estimating the Benefits of Visibility Improvement in the Northeast" US EPA, WAshington,D.C. 1984.

Tukey,John, Experimental Data Analysis Reading,MA: Addison-Wesley, 1970.